GNR 638 (Spring 2024): Mini Project 1
Fine Grained Classification on CUB Dataset Using CNN

Soumen Mondal (23m2157)
Siddhant Gole (23m2154)
Akash Pal (23m2158)

March 5, 2024

1 Introduction

This project focuses on addressing the challenging task of fine-grained classification in computer vision, where objects are categorized into highly specialized classes with subtle differences. The motivation stems from the importance of fine-grained classification in real-world applications and the need for efficient models to tackle this task. The primary objectives include training a Convolutional Neural Network (CNN) model with a maximum of 10 million parameters, achieving high accuracy, and ensuring parameter and training time efficiency. The scope of the report encompasses a detailed methodology, experimental setup, results, discussion, and conclusion, providing insights into the effectiveness of the proposed approach in addressing fine-grained classification challenges.

2 Methodology

2.1 Dataset

The dataset used for training and evaluation is the CUB (Caltech-UCSD Birds-200-2011) dataset, a widely-used benchmark for fine-grained classification tasks in computer vision. It contains images of 200 bird species, with each species having a varying number of images. The dataset provides annotations for bounding boxes around the birds, enabling precise localization of the objects. Additionally, it includes attributes for each bird species, which can further aid in classification tasks by providing additional semantic information.

2.2 Model Selection

The choice of the EfficientNet-B0 architecture for fine-grained classification is motivated by its superior performance in terms of parameter efficiency and accuracy. EfficientNet models are specifically designed to achieve better trade-offs between accuracy and computational resources by scaling the network architecture in a principled manner. The EfficientNet-B0 variant, being the smallest and least computationally intensive model in the EfficientNet family, strikes a balance between model complexity and performance, making it well-suited for fine-grained classification tasks. Its efficient use of parameters allows for faster training and inference without compromising on accuracy, making it an ideal choice for resource-constrained environments.

2.3 Transfer Learning

Transfer learning from ImageNet-pretrained weights was employed to initialize the model's weights and improve convergence during training. By leveraging the pre-trained weights from a large-scale dataset like ImageNet, which contains millions of images across thousands of categories, the model can benefit from learning generic features that are transferable to the fine-grained classification task. This initialization helps the model to start from a better initialization point, allowing it to learn task-specific features more efficiently and effectively. Fine-tuning the pre-trained model on the target dataset further refines the learned representations, adapting them to the specific characteristics of the CUB dataset and improving overall classification performance.

3 Experimental Setup

3.1 Training Configuration

3.1 Training Configuration Description of the training parameters, including batch size, learning rate, optimizer, and any regularization techniques employed.
3.2 Evaluation Metrics Explanation of the evaluation metrics used to assess the model's performance, such as accuracy, precision, recall, and F1-score.

4 Results

4.1 Training Process Visualization of the training loss and accuracy curves over epochs to demonstrate the model's convergence and performance during training.
4.2 Performance Evaluation Presentation of the final accuracy achieved on the test set and comparison with baseline models or state-of-the-art approaches.

5 Discussion

5.1 Interpretation of Results Analysis of the findings, including insights into the model's performance, strengths, limitations, and areas for improvement.
5.2 Parameter Efficiency Discussion on how the model's parameter efficiency contributed to its effectiveness in fine-grained classification tasks.
5.3 Training Time Efficiency Evaluation of the training time efficiency in terms of the number of iterations required to achieve convergence.

6 Conclusion

6.1 Summary of Findings A summary of the key findings and contributions of the project, highlighting the effectiveness of the EfficientNet-B0 model for fine-grained classification.
6.2 Implications and Future Work Discussion on the implications of the findings and suggestions for future research directions, such as exploring other efficient architectures or datasets.
6.3 Closing Remarks Final remarks on the significance of the project and its potential impact on the field of computer vision.